Clustering Results Report

This report outlines the results of clustering analysis performed on the E-Commerce dataset. The goal was to segment data points into distinct clusters based on their features. This report summarizes the number of clusters formed, evaluation metrics, and provides visualizations to validate the clustering results.

- I have used k-means clustering algorithm on the given dataset. K-Means is an intuitive algorithm that groups data based on proximity to cluster centroids. Its ease of implementation and speed makes it suitable for this dataset. The clustering algorithm segmented the data into 5 clusters. These clusters represent distinct groups based on the similarity of data points.
  Number of clusters are 5 (k=5).

Davies-Bouldin Index

The Davies-Bouldin Index (DB Index) measures the average similarity ratio of each cluster with its most similar cluster. A lower DB Index indicates better-defined clusters.

DB Index Value: 0.7684

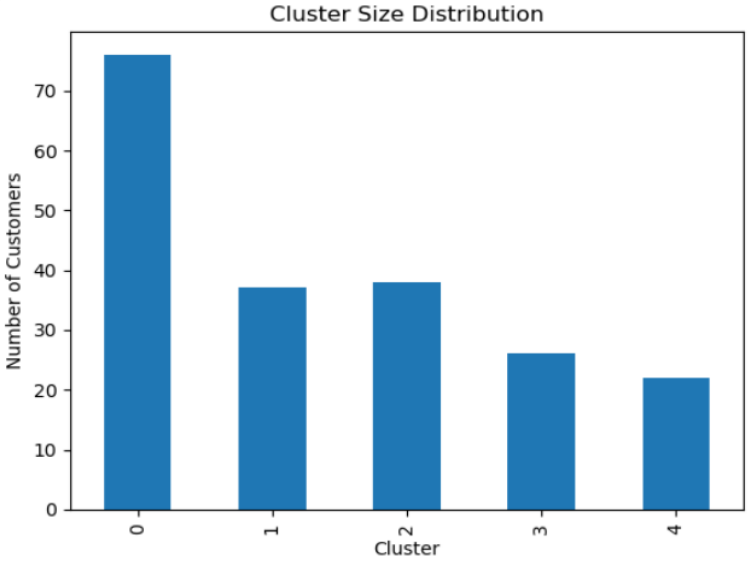
- As the DB index value is 0.7684, suggests that the clusters are moderately compact and fairly well-separated. This value indicates good clustering performance in scenarios where clusters are not expected to have significant overlaps.

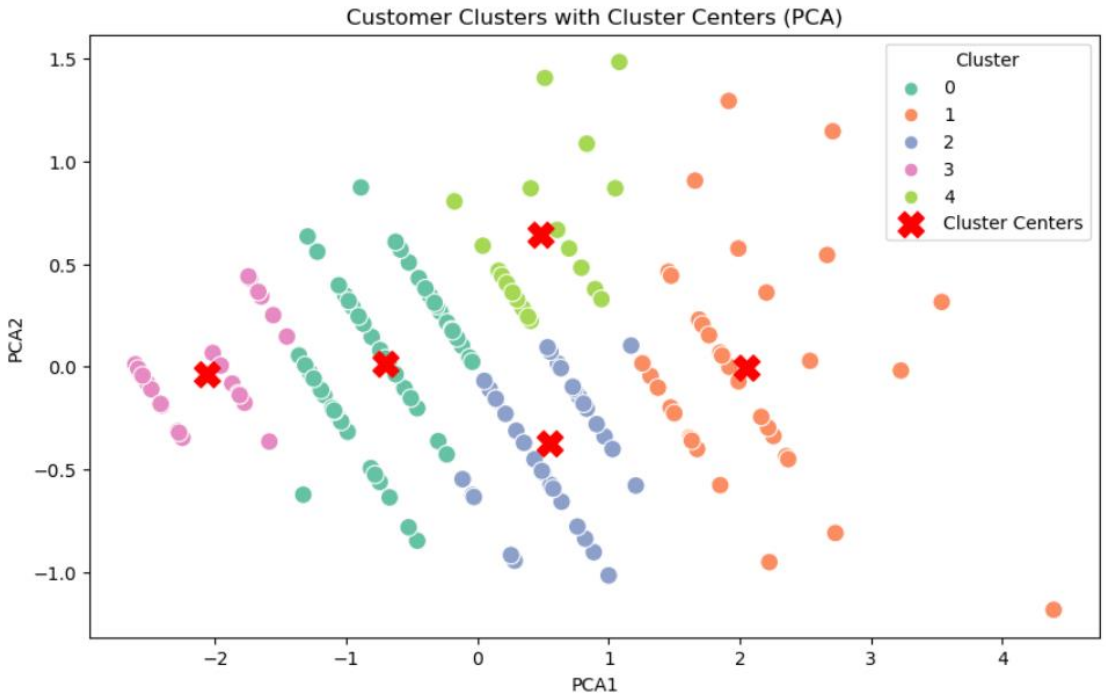
Silhouette Score

The Silhouette Score evaluates how similar a data point is to its own cluster compared to others. It ranges from -1 to 1, where a higher score indicates well-separated clusters.

Silhouette Score: 0.4026

- A score of 0.4026 indicates moderate clustering quality. Data points are more closely aligned with their own clusters than with others.





- The clustering analysis successfully segmented the dataset into 5 meaningful clusters. The evaluation metrics indicate that the clustering algorithm provided moderate performance.

- PCA visualization further validates the distinctiveness of the clusters. These results can be used for targeted strategies and decision-making processes.